European Online Journal of Natural and Social Sciences 2013;
Vol.2, No.3 Special Issue on Accounting and Management.
ISSN 1805-3602

www.european-science.com

Financial Distress Prediction of Tehran Stock Exchange Companies Using Support Vector Machine

Afshin Niknya[1]*, Roya Darabi[2], Hamid Reza Vakili Fard[1]

[1]Accounting and Management Department, Science & Research Branch , Islamic Azad University, Tehran, Iran

*E-Mail: afshin.niknya@gmail.com

[2]Department of Accounting and Management, South Tehran Branch, Islamic Azad University, Tehran, Iran

Abstract

The main objective of this study is to evaluate and to compare the power to predict company financial distress by utilizing the Support Vector Machine (SVM) to the multiple-discriminant analysis and the logistic regression models. Companies approved for acceptance into Tehran Stock Exchange Market between 2007 and 2013 comprise the statistical population for the study. In order to predict financial distress based on financial ratios such as profitability, activity ratio, ratios per share, etc. by using the Support Vector Machine (SVM), the sample data has been divided into two separate groups: the training group and the experimental group. The training set is made up of 540 year-company and the experimental set is comprised of 120 companies in 2013. Finally, conclusions obtained from SVM, multiple-discriminant analysis and the logistic regression models for predicting financial failure were surveyed and compared. Results of testing hypothesis indicate with a 95% certainty ratio that there is a significant difference in the average prediction accuracy of the three models. Consequently among the three, the SVM model has the highest accuracy level for predicting company financial failure and the multiple-discriminant analysis model has the lowest.

Keywords: Financial failure, Support Vector Machine (SVM), multiple discriminant analysis, logistic regression model

Introduction

Development of new technologies and their application in various fields of science have attracted the attention of accounting professionals and those engaged in the area of financial management. This interest has ultimately led to the utilization of such innovations in these two sectors. Transformations triggered by the innovative technologies and their wide range of scientific applications has persuaded analysts and accountants alike to use them in order to boost the effectiveness of their decision making. Considerable research has been carried out globally to develop and to provide models which have the power to predict company financial distress. Due to this fact, an ever increasing interest in theoretical development of dynamic intelligent systems based on empirical data has been gaining momentum. The Support Vector Machine (SVM) is one of these intelligent systems which through processing sets of empirical data can decipher the hidden codes lying within transforming them into algorithm and network structures. There are numerous examples of companies in the domain of accounting and financial management which have experienced financial distress in the past and there are also many which have not. Characteristics of these companies along with their existing accounting data are a valuable source to be applied as algorithm input for problems pertaining to the prediction of financial failure and the resulting output can be used to predict company financial failure.

The main issue for this study is to review and to compare the power to predict company financial failure by utilizing the Support Vector Machine (SVM) versus the multiple discriminant analysis and the logistic regression models.

Objectives of this current study include the following:

- To predict financial distress for companies which are considered for the study by utilizing the Support Vector Machine (SVM) and the multiple discriminant analysis and the logistic regression models and to finally evaluate the efficiency level concerning those very models, and
- To compare the power of such models to predict financial distress for companies considered for the study.

Theoretical Literature and Study Background

Lars Chamberlain (1911) in an article titled “Principles of Bond Investment” was able to achieve performance ratios from ratios yielded by Woodlock.

Arthur Venekor and Raymond Smith (1935) concluded in their study titled “Methods in the Analysis of Financial Ratios of Bankrupt Companies” that the most accurate ratio for determining the status of bankruptcy is the ratio of working capital to total asset.

Beaver (1966) selected a set including thirty financial ratios which in his opinion were the most suitable for evaluating the well-being state of a company. He then classified the ratios according to company evaluation and concluded that the value for every single ratio is directly related to the classification of a company as bankrupt or non-bankrupt in which low classification of error margin signifies higher value of each ratio. According to this principle, Beaver introduced the ratio which retained the lowest classification error margin in order of importance as follows: cash flow to total assets, net income to total assets, total of debt to total assets, working capital to total assets, current ratio and ratio of distance uncertainty.

After initial studies conducted by White in 1988, the door for artificial neural networks to move into the financial sector was opened leading to numerous studies conducted globally. From 1995 to 1998 a total of 213 different activities were carried out in the commerce sector, 54 of which were in the financial domain and 2 were in the domain of prediction and analysis of time series (Woong et al. 1977, Sinaie 2006).

Zhiang et al. (1996) used genetic algorithm to project net asset prices for investment companies at the end of their fiscal year. They compared the network data and their conclusions with results obtained from the more traditional techniques used to measure state of economy and discovered that if data input is low then the genetic algorithm has a significantly better performance than the regression methods. In a separate study they tested their model on the Taiwan market by developing a genetic algorithm which allowed for the input of political as well as quantitative factors.

Aiken and Basset (1999) used a forward neural network which had been trained using the genetic algorithm method to project the interest rate of the US Treasury and concluded that the neural network can be appropriate for this purpose.

Garliaks (1999) used the genetic algorithm associated with a kernel function and the back-propagation prediction method to project the time series of the stock market. Based on his conclusions, projecting financial cycles is carried out better using the genetic algorithm as compared to the classic statistical models or other similar models.

Chan et al. (2000) also applied the genetic algorithm along with data obtained from daily transactions at Shanghai Stock Exchange in order to project financial cycles. For higher speed and consistency, they used the gradient descent algorithm and the multiple linear regressions to determine weights. They concluded that the genetic algorithm can make more satisfactory projections on time cycles and also the weight method they opted to use required less time in performing calculations.

Kim and Han (2000) applied an adjusted neural network by genetic algorithm to project stock index. In this case the genetic algorithm was utilized to reduce the complexity of future price cycles.

Carlos Batista (2000) used the genetic algorithm in his attempts to project the stock price index at the Philippines Stock Exchange and concluded that there is no significance difference between applying the genetic algorithm or the random walk hypothesis to short interruptions; however, for long interruptions the genetic algorithm is more effective in projecting indices.

Landas et al. (2000) also attempted to project indices using the genetic algorithm. Their data were divided into two different types: internal and external. External data included TOPIX stock price indices, SBF250 and S & P 500, exchange rates for dollar.mark and dollar.yen as well as the three-month interest rate and the Treasury interest rate. Internal data only included the number of indices. Their conclusion indicated that genetic algorithm has a better performance than linear methods.

Olson and Mossman (2003) through using the relevant data from 2352 Canadian companies from 1976 to 1993 were able to predict stock return. They used the three methods of genetic algorithm, ordinary least squares and logistic regression to project the rate of return. Current ratio, liquidity ratio, inventory turnover ratio, debt to equity ratio, return to equity ratio, ratio of book value to market value and price to sales of variables were considered in their study. The conclusion of their study also indicated that genetic algorithm is superior to the other two methods.

Pav (2008) used the relevant data of 720 Taiwanese companies from 2000 to 2005 to compare the genetic algorithm method to multiple regressions in connection to modelling the capital structure. Conclusions showed that the model provided by the genetic algorithm had a lesser error margin than the multiple regressions model.

Henry Hoglond (2012) in his study under the title “Discovering Earnings Management through the Utilization of Neural Network” compared the capability of the regression model to the neural network model. The results of his study indicated that the neural model network is more accurate in the discovery of earnings management as compared to the regression model and as result it can be considered a more feasible means.

Campello et al. (2013) reviewed the impacts of various organizational factors on financial crisis experienced by companies. In this particular study, capital expenditure, marketing and sales expenditure, number of personnel, sales growth, profitability and company size were considered as independent factors so their effect on company financial crisis can be examined. They applied multivariate regression in their study based on time cycle data and concluded that capital expenditure, marketing and sales expenditure, number of personnel, sales growth, profitability and company size have a direct impact on financial crisis experienced by companies.

Henry Hoglond (2013) in his study under the title “Bankruptcy Prediction Using Support Vector Machine” compared the power of this model to logistic regression and multiple discriminant analysis models. In order to predict financial failure using SVM based on financial ratios such as profitability, activity ratios, ratios per share, etc. the sample data was divided into two groups: training and experimental for which the testing hypothesis conclusions indicate with a 95% certainty ratio that a significant difference exists in the average prediction accuracy of the three models. This literally means that the SVM model has the highest accuracy level and the multiple discriminant analysis model the lowest in predicting company financial failure.

Hypotheses of the study

H1: SVM model has a higher power in predicting company financial failure as compared to the linear regression model.

H2: SVM model has a higher power in predicting company financial failure as compared to the multiple-discriminant analysis model.

H3: The model of multiple discriminant analysis has a higher power in predicting company financial failure as compared to the linear regression model.

Design and Methodology

This study falls in the area of positive research with the main objective of applicability and since data from past studies were used in order to test the study hypotheses, it may be categorized as a quasi-experiment. This current study is of inductive and casual type in which the Support Vector Machine (SVM) is utilized in order to project the multiple discriminant analysis and the logistic regression of financial failure for companies and to compare the power of these three methods to make appropriate projections.

Methods of collecting data applied in conducting this study are library research, relevant articles, domestic and international dissertations as well as information provided by the Stock Exchange Market. The location for conducting the study is Tehran Stock Exchange. Since the time period for the study is from 2007 to the end of 2013, the statistical population includes all companies approved and accepted by Tehran Stock Exchange. Sampling method is systematic elimination which is applied based on the following prerequisites:

- The required data for calculating operational variables of those companies must be accessible. They have to have been accepted in the Stock Exchange Market at least since 2007 and must remain active in same all the way through to the conclusion of the study. The end of fiscal year for these companies must be March, 19, 2013 and they must not be considered as financial and investment institutes or banks.

The dependent variable of this study is financial distress. If under Article 141 of Trade Code a company is considered to be experiencing financial failure, then number one is applied, otherwise; zero is used. After a thorough study of the research literature was conducted, thirty independent variables in six different categories - which were utilized more frequently in separate studies carried out previously - were selected as follows:

Profitability ratios including: Gross profit on sales, net profit on sales, earnings before interest, tax to total assets, net income to total assets, net income to total current assets, net income to total fixed assets, gross profit margin, the aggregate net profit of equity

Activity ratios including: Accounts receivable turnover, inventory turnover, accounts payable turnover, circulating assets, circulating current assets, circulating fixed assets

Debt ratios including: Current ratio, ratio of debt to cash flow from operating activities, ratio of debt to total assets, ratio of debt to tangible fixed assets, ratio of debt to market value equity, debt to equity ratio, interest expense coverage ratio

*Growth ratios:*Asset growth rate, sales growth rate

Structural Ratios: Ratio of current assets to total assets, ratio of fixed assets to total assets, ratio of equity to fixed assets, ratio of current liabilities to total of debts

Ratios per share: Earnings per share, net assets per share, cash yield from operating activities of a single share

Methods and tools for data analysis

Since the main objective of this study is to analyse and to compare the power to predict financial failure for companies accepted in Tehran Stock Exchange Market using a Support Vector Machine (SVM) to multiple discriminant analysis and logistic regression models, accordingly financial distress of companies has been calculated using these three methods as well as the data obtained from Stock Exchange online site and also through the application of innovative and smart EViews and Matlab software. Moreover, the margin of error has also been compared for these three methods and is utilized to review and to test the study hypotheses.

Empirical Results of the Study

Table 1. Descriptive Statistics for Study Variables

Maximum	Minimum	Standard Deviation	Median	Mean	Description of Variables		
0.524	0.005	0.403	0.138	0.0791	X_1	Gross Profit on Sales	Profitability Ratios
0.522	0.003	0.400	0.115	0.085	X_2	Net Profit to Sales	
0.113	0.026	0.069	0.112	0.053	X_3	Earnings before Interest	
0.745	0.000	1.112	0.167	0.121	X_4	Net Income to Total Assets	
0.720	1.010	0.600	0.863	1.259	X_5	Net Income to Total Current Assets	
0.524	0.005	0.403	0.138	0.080	X_6	Net Sum of Fixed	
0.522	0.003	0.400	0.115	0.085	X_7	Gross Profit Margin	
1.259	0.000	0.053	0.085	0.080	X_8	Aggregate Net Profit of Equity	
0.863	0.000	0.112	0.115	0.138	X_9	Accounts Receivable Turnover	Activity Ratios
0.600	0.200	0.069	0.400	0.403	X_{10}	Inventory Turnover	
1.010	0.000	0.026	0.003	0.005	X_{11}	Accounts Payable Turnover	
0.720	0.330	0.113	0.522	0.524	X_{12}	Circulating Assets	
19.618	9.536	1.342	12.659	12.822	X_{13}	Circulating Current	
0.726	0.000	0.107	0.105	0.133	X_{14}	Circulating Fixed Assets	
0.152	0.001	0.042	0.079	0.078	X_{15}	Current Ratio	Debt Ratios
0.777	0.002	0.136	0.097	0.138	X_{16}	Ratio of Debt to Cash Flow from Operating	
0.600	0.200	0.138	0.400	0.389	X_{17}	Ratio of Debt to Total Assets	
0.018	0.000	0.003	0.003	0.004	X_{18}	Ratio of Debt to Tangible Fixed Assets	
0.335117	0.121724	0.080784	0.095979	-0.18	X_{10}	Ratio of Debt to Market Value of Equity	
0.725	1103.056	1450.000	1598.124	2856.503	X_{20}	Debt to Equity Ratio	
0.224217	0.155391	0.4	0.374599	0.60	X_{21}	Interest Expense Coverage Ratio	
-0.26804	0.340388	0.6025	0.514711	0.54	X_{22}	Asset Growth Rate	Growth Rates
0.029851	0.180659	0.555621	0.55049	0.84	X_{23}	Rate of Sales Growth	

1.832105	0.107449	0.118604	0.136022	0.25	X_{24}	Ratio of Current Assets to Total Assets	Structural Ratios
0.788769	1.400664	13.03148	13.17033	13.07	X_{25}	Ratio of Fixed Assets to Total Assets	
0.02365	2.300908	7.682836	7.600146	6.00	X_{26}	Ratio of Equity to Fixed Assets	
0.335117	0.121724	0.080784	0.095979	0.18	X_{27}	Ratio of Current Liabilities to Collect	
0.725	1103.056	1450.000	1598.124	2856.503	X_{28}	Earnings per Share	Ratios per Share
0.224217	0.155391	0.4	0.374599	0.60	X_{29}	Net Assets per Share	
0.679	0.232	1.403	0.332	0.453	X_{30}	Cash from Operating Activities per Share	

Descriptive Statistics

Table 1 represents the descriptive statistics for study variables for the duration of the research. After required modifications in order to identify companies which lack proper qualifications and also to eliminate unnecessary data, the total for findings equals 110 companies (660 year-company). Descriptive statistics for dependent and independent variables which were measured using data obtained from 110 companies (660 year….company) from the testing period of 2007 to 2013 include mean, median, standard deviation, minimum and maximum which are represented in Chart 1 below.

Designating an Appropriate Model to Estimate the Regression Model

As pointed out in Chapter III in cases when correlation of a dependent variable and one or more independent variables is considered and the researcher aims to estimate parameter(s) for the dependent variable(s) based on this notion and application of historical data to make a prediction, the existing data and a variables in a model are usually three different types including Time Series Data, Cross Section Data, and Pooling Data

Time series data measures the values of a variable(s) at consecutive points in time. This sequence may be annual, quarterly, monthly, weekly or it may even be continuous.

Cross section data measures the values of a variable(s) over time and over multiple units. Such units may be manufacturing and industrial units or may be different companies.

Pooling data is literally cross section data over length of time or another word such data is yielded through the integration of time series data and cross section data.

Due to the study literature involved and also the nature of the research hypotheses, pooling data has been utilized in carrying out this study. Moreover, in order to designate an appropriate model (consolidated or panel with fixed or random effects) for testing the study hypotheses, the Chav and Hausman tests have been used.

A) Chav Test

Results for test F for the study regression model are shown in Chart 2, for this model the Chav Test indicates the non-confirmation of the H0 hypothesis (consolidated model). Another word, there are individual or group effects; therefore, the panel method has to be used to estimate the study regression model. Accordingly, the Hausman Test is applied in order to designate the type of panel model (fixed or random effects).

Table 2. Chav Test

Test Result	Probability	Statistics F	Test F	Description
-	0.0000	83.021**	Value	Regression Model

**Significant at 95%

B) Hausman Test

After determining that the intercept is not consistent for different years, the method for estimating the model (fixed or random effects) must be designated using the Hausman Test**:**

In Hausman Test, the H_0 hypothesis is tested based on the estimate compatibility produced for random effects versus the H_1 hypothesis based on the estimate incompatibility produced for random effects.

Table 3 . Hausman Test

Test Result	Probability	Statistics χ^2	Hausman Test	Description
Hypothesis Rejection	0.0086	47.091**	Value	Study Regression Model

**Significant at 95%

The results of Hausman Test are shown in Chart 3. As indicated by the results the χ^2 statistics of the Hausman Test is equal to 47.091 which is significant at 95% certainty level. Thus, according to this test, fitting the regression model in this study shall be appropriate only if panel model data is applied as fixed effect method.

Testing the Classic Regression Hypotheses

Prior to fitting the regression models in the study, the linear regression hypotheses must be put to test.

Testing the Normal Distribution of Study Variables

In order to test the normal distribution of variables related to the study, the Kolmogorov-Smirnov Test has been utilized. This test has been conducted for the dependent variable (financial failure). Test output table K-S in SPSS software for this variable has been outlined in table 4:

Table 4. Kolmogorov-Smirnov Test

Result	Significance Level	Kolmogorov-Smirnov Z	Variable Name
Distribution is Normal	0.124	1.328	Financial Failure

With regards to the above chart and the Kolmogorov-Smirnov Z statistics which indicate a higher significance level than 0.05 for all variables, the H0 hypothesis is confirmed. Accordingly, it can be deduced with a 95% certainty that the distribution of above variables is normal.

Testing the Independence of Errors

The Durbin-Watson Test examines the serial correlation of the remaining regression error(s) based on the statistical zero hypothesis outlined below:

H0: There is no correlation among errors.

H1: There is correlation among errors.

If the statistics for Durbin-Watson Test fall between 1.5 and 2.5 then the H0 test hypothesis is accepted (non-correlation among errors); otherwise, H1 is confirmed.

Durbin-Watson statistics together with correlation coefficient, coefficient of determination, adjusted coefficient of determination and standard error are shown in Chart 5:

Table 5. Error Independence Test

Durbin-Watson Statistics	Adjusted Coefficient Determination	Coefficient of Determination	Description
1.942	2.70	8.76	Regression Model

According to the above table, the value for Durbin-Watson statistics for the regression models are between 1.5 and 2.5; therefore, the H0 hypothesis based on lack of correlation among errors is confirmed and the regression may be used.

Examining the Normal Distribution of Errors

One hypothesis considered for regression is that equation errors have a normal distribution with a mean of zero. In order to examine the normal state of equation errors, a curve showing error components in regression model is drawn – as shown in Figure 1. In the study regression model error distribution is nearly zero and its standard deviation is close to one (0.994). As result distribution of errors for the regression model is normal.

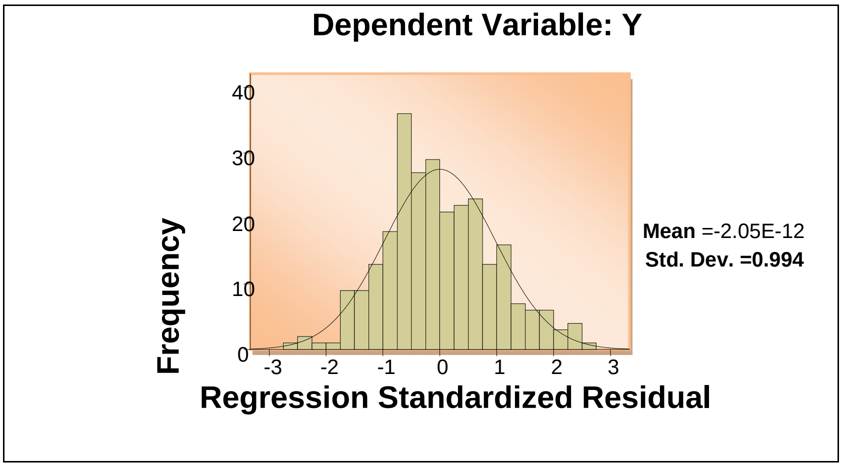

Figure 1. Error Component Curve – Model(1)

Multiple Diagnostic Analysis (MDA)

In order to predict financial failure and to separate companies experiencing such failure from others, the study independent variables have been used in six financial ratio groups including: profitability ratios, activity ratios, debt ratios, growth ratios, structural ratios and ratios per share. This has been carried out through maximizing the variance among groups based on a statistical decision rule for which the linear combination is as follows:

$$z = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$$

In this equation Z is separation of companies experiencing financial failure from others, W is detection weight and X represents independent variables including the six aforementioned ratios.

Table 6. Detection Weight for Each of the Six Ratios

Weight (W)	Variable Description		
7365.0	X1	Gross Profit on Sales	Profitability Ratio
-0.6656	X2	Net Profit on Sales	
0.5053	X3	Earnings before Interest and Tax to Total	
-0.3892	X4	Net Income to Total Assets	
0.4845	X5	Net Income to Total Current Assets	
-0.4528	X6	Net Sum of Fixed Assets	
0.3770	X7	Gross Profit Margin	
-0.5421	X8	Aggregate Net Profit of Equity	
0.4548	X9	Accounts Receivable Turnover	Activity Ratio
-0.5162	X10	Inventory Turnover	
-0.5981	X11	Accounts Payable Turnover	
-0.5146	X12	Circulating Assets	
0.5331	X13	Circulating Current Assets	
-0.4091	X14	Circulating Fixed Assets	
0.4620	X15	Current Ratio	Debt Ratio
-0.4900	X16	Ratio of Debt to Cash Flow from Operating Activities	
0.6658	X17	Ratio of Debt to Total Assets	
-0.4185	X18	Ratio of Debt to Tangible Fixed Assets	
0.7118	X19	Ratio of Debt to Market Value of Equity	
-0.4817	X20	Debt to Equity Ratio	
0.6702	X21	Interest Expense Coverage Ratio	
-0.6266	X22	Asset Growth Rate	Growth Rates
-0.4338	X23	Rate of Sales Growth	
-0.6881	X24	Ratio of Current Assets to Total Assets	Structural Ratios
0.5778	X25	Ratio of Fixed Assets to Total Assets	
-0.6782	X26	Ratio of Equity to Fixed Assets	
0.5243	X27	Ratio of Current Liabilities to Collect Debts	
-0.7345	X28	Earnings per Share	Ratios per Share
0.5158	X29	Net Assets per Share	
-0.4768	X30	Cash from Operating Activities per Share	

Support Vector Machine (SVM)

The support Vector Machine (SVM) has been used in this study in order to predict financial distress. Furthermore, Matlab software has been used to design the SVM. The set of data considered for this purpose include information such as profitability, activity ratio, shares ratios, etc. for 110 companies (660 year-company) from the start of 2007 to the end of 2013. Sample data is also divided into two groups: training and experimental. The training set used for developing the model includes 540 year-company from 2007 to 2013 while the experimental set which includes 120 companies is only used to evaluate the validity and interoperability of the applied model. Training data is required for SVM design. The number of times which the training stage is repeated is presumed to be 1000 system.

Results of the Logistic Regression

After testing regression hypotheses and gaining enough assurance of their development, the yielded results for applying the above regression equation has been presented in table 7. Statistics value of 8.654 indicates the significance of the regression model. As it can be seen at the bottom of the chart, the coefficient of determination and the adjusted coefficient of determination 8.76% and 2.70%, respectively so it can be concluded that in this regression equation only around 2.70% of financial failure of companies considered for the study may be explained by independent variables.

In this table, positive (negative) numbers in the ratio value column indicate direct effect (reverse) of each variable on financial failure of companies considered for the study.

Method of Judgement: If the sig value calculated by the software is less than the certainty level considered (equal to 5% in this study), the significance of the variable taken into account is then confirmed. Also according to the Wald Statistics if this value with the same certainty level (5%) is more than its equivalent value in the Wald Student Chart then the significance of the variable is confirmed. Results yielded from significance of variable ratios indicate whether a given variable will suffice in predicting financial failure or not.

Table 7. Results from Logistic Regression Equation

Significance Level	Wald Statistics	Coefficient Ratio	Differential	Variable	
0.002	3.154	1.561	a0	Constant	
0.018	2.381	1.421	a1	X1	Gross Profit on Sales
0.009	2.619	1.527	a2	X2	Net Profit on Sales
0.046	-2.002	-0.376	a3	X3	Earnings before Interest and Tax to Total Assets
0.823	-2.223	-0.651	a4	X4	Net Income to Total Assets
0.015	2.451	0.452	a5	X5	Net Income to Total Current Assets
0.014	2.479	0.781	a6	X6	Net Sum of Fixed Assets
0.000	5.073	1.034	a7	X7	Gross Profit Margin
0.000	3.838	1.081	a8	X8	Aggregate Net Profit of Equity
0.002	2.388	0.631	a9	X9	Accounts Receivable Turnover
0.003	-2.141	-0.753	a10	X10	Inventory Turnover
0.0037	-2.601	-2.894	a11	X11	Accounts Payable Turnover

0.541	0.671	0.711	a12	X12	Circulating Assets
0.121	0.576	0.967	a13	X13	Circulating Current Assets
0.034	2.311	0.892	a14	X14	Circulating Fixed Assets
0.0067	3.073	0.453	a15	X15	Current Ratio
0.027	2.673	1.235	a16	X16	Ratio of Debt to Cash Flow from Operating
0.035	2.388	0.947	a17	X17	Ratio of Debt to Total Assets
0.043	-2.141	-1.236	a18	X18	Ratio of Debt to Tangible Fixed Assets
0.058	1.052	2.034	a19	X19	Ratio of Debt to Market Value of Equity
0.069	3.838	3.467	a20	X20	Debt to Equity Ratio
0.321	0.897	1.091	a21	X21	Interest Expense Coverage Ratio
0.281	-0.458	-0.818	a22	X22	Asset Growth Rate
0.0037	2.987	1.501	a23	X23	Rate of Sales Growth
0.014	-2.847	-0.818	a24	X24	Ratio of Current Assets to Total Assets
0.004	2.873	3.641	a25	X25	Ratio of Fixed Assets to Total Assets
0.000	5.073	2.589	a26	X26	Ratio of Equity to Fixed Assets
0.000	3.838	3.098	a27	X27	Ratio of Current Liabilities to Collect Debts
0.002	2.388	1.098	a28	X28	Earnings per Share
0.003	-2.141	-3.215	a29	X29	Net Assets per Share
0.001	-2.726	-1.982	a30	X30	Cash from Operating Activities per Share
8.654	X2		8.76	Coefficient of Determination	
0.017	Significance (P-Value)		2.70	Adjusted Coefficient of Determination	
1.942	Durbin-Watson Statistics				

Results for the Support Vector Machine (SVM)

MSE (Mean Square Deviation) Graph

As shown by the following graph, the MSE for training data is 10-20 and for testing data it equals 10-2 repeated 1,000 times which is considered appropriate.

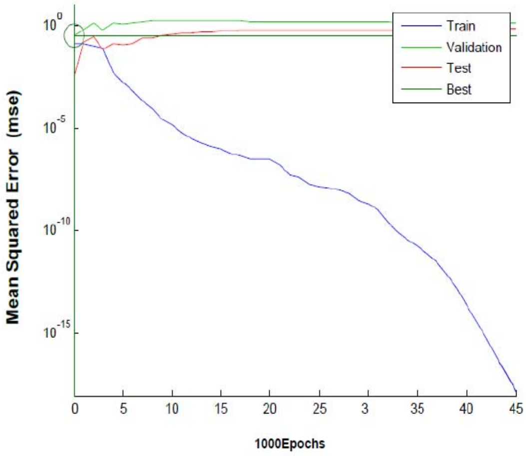

Figure 2. MSE Changes

Surveying Data Prediction Error

As it is evident from figure 3, errors due to predicting financial failure among the statistical population are negligible and can be ignored which in a way is a testament to the high capability of the designed SVM in making accurate predictions.

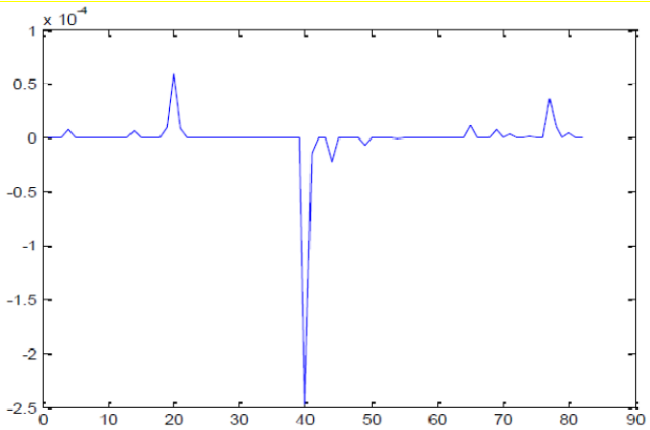

Figure 3. Data Prediction Error

Surveying and Comparing the Prediction Power of Logistic Regression, SVM and Multiple Discriminant Analysis

Power of the three models, namely logistic regression, SVM and multiple discriminant analysis to predict companies which are experiencing financial failure and those companies lacking such failure have been presented in Table 8. The first row in each group shows the number of correct predictions for 60 actual cases. To compare power of the three models put forth here in predicting financial failure or lack thereof, test of mean comparison of communities (Statistics F) has been used. Results of this test are shown in Table 9. In Test F, hypotheses H0 and H1 are as follows:

Table 8. Comparison of Prediction Capability for Different Models

Multiple Discriminant Analysis	SVM	Logistic Regression	Group Name
47	57	50	Financial Failure
%78.33	%95.00	%83.33	
51	55	52	Financial Non-Failure
%85.00	%91.66	%86.66	
98	112	102	Overall Model Accuracy
%81.66	% 93.33	%85.00	

H0: There is no significant difference in mean accuracy of logistic regression, SVM and multiple discriminant analysis models.

H1: There is a significant difference in mean accuracy of logistic regression, SVM and multiple discriminant analysis models.

Table 9. Mean accuracy for Model Prediction

Multiple Discriminant Analysis	SVM	Logistic Regression	Method
%81.66	% 93.33	%85.00	Overall Mean accuracy
		9.451	Statistics F
		0.0021	P-) Significance (Value

Results for Test F on comparison of prediction mean accuracy of the three models have been presented in accordance with Table 9 indicating a significant difference in the prediction mean accuracy of the three models at a 95% certainty level. This difference is especially due to the fact that the value for Statistics F is 9.4516 more than what is acceptable with a 95% certainty level. Consequently, at the acceptable error level of 5% the statistical hypothesis for significance of the difference in prediction mean accuracy of the three models is not rejected and the H1 hypothesis indicating a significant difference among logistic regression, SVM and multiple-discriminant analysis models is confirmed.

Results of Hypotheses Testing

The First Hypothesis**:** The SVM model has a higher level of capability to predict company financial failure than the linear regression model.

Results for Test F on comparison of prediction mean accuracy for the three models have been presented in accordance with Table 9 indicating a significant difference in the average prediction accuracy of the three models at a 95% certainty level. Accordingly, since the accuracy level in the SVM model (93.33%) is more than the linear regression model (85.00%), the first study hypothesis that the SVM model possesses a higher level of power in predicting company financial failure as compared to the linear regression model is confirmed.

The Second Hypothesis: The SVM model has a higher level of power to predict company financial failure than the multiple-discriminant analysis model.

Results for Test F on comparison of prediction mean accuracy for the three models have been presented in accordance with Table 9 indicating a significant difference in the prediction mean accuracy of the three models at a 95% certainty level. Accordingly, since the accuracy level in the SVM model (93.33%) is more than the multiple discriminant analysis model (81.66%), the second study hypothesis that the SVM model possesses a higher level of power in predicting company financial failure as compared to the linear regression model is confirmed.

The Third Hypothesis: The multiple-discriminant analysis model has a higher level of power to predict company financial failure than the linear regression model.

Results for Test F on comparison of average prediction accuracy for the three models have been presented in accordance with Table 9 indicating a significant difference in the prediction mean accuracy of the three models at a 95% certainty level. Accordingly, since the accuracy level in the linear regression model (85.00%) is more than the multiple- discriminant analysis model (81.66%), the third study hypothesis that the multiple-discriminant analysis model possesses a higher level of capability in predicting company financial failure as compared to the linear regression model is rejected.

Conclusion

One of the ways we can use for proper utilization of investment opportunities and better allocation of resources is the prediction of financial failures. First, such predictions provide essential warnings for companies of an impending failure and enable them to prepare accordingly. Second, predicting financial failures encourages investors and creditors to distinguish suitable and ideal investment opportunities from the more undesirable ones and channel their investments into the right trajectory (Ra'ai and Falalhpoor, 2008). On the other hand technological progress and innovation along with its application in various fields of science have most certainly attracted the attention of those involved in accounting and financial management and have paved the way for the utilization of such processes in both those two sectors. Transformation of technology and its wide spread scientific use has led to an increase in its related applications in the field of accounting and today an ever increasing number of accountants are using technology in order to boost their efficiency and productivity. One of the most important ways to increase accuracy in predicting company financial distress is the application of data mining when making predictions. A considerable number of international researches have been carried out so far to develop and to introduce models which can readily identify companies that are at risk of financial distress. They have sparked interest in theoretical development of dynamic intelligent systems following these models and based on empirical data. The Support Vector Machine (SVM) is one of these intelligent systems which through processing sets of empirical data can decipher the hidden codes lying within transforming them into algorithm and network structures.

With full regards to the aforementioned facts, it can be deduced that the main issue of concern to this study is to survey and to compare the level of capability to predict company financial distress by utilizing the Support Vector Machine (SVM) versus the logistic regression model. Accordingly, 660 year-company were selected as sample, 112 of which were financially distressed and 548 of which were non-distressed. In order to ensure the validity of the model, the sample was divided into two samples: the training sample and the experimental sample

References

Aziz, A.E, Diar, L, (2006) "Bankruptcy Predictio, Where We Stand?" Journal of Corporate Governance, Vol 6, No. 1.

Altman, E.I, Haldeman, R.G, Narayanan, P, (1977) "ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations" Journal of Banking and Finance 1 (1).
Altman, Edward, I, (1968) "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy" The Journal of Finance 23.
Atiya, A. (2001). Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results. IEEE Transactions of Neural Networks, 12(4).
Bäck, T (1996) “Evolutionary Algorithms in Theory and Practice” ISBN:0-19-509971-0.
Beaver, W.H. (1966) “Financial Ratios as Predictors of Failures, In Empirical Research in Accounting”, Selected Studies Supplement to the Journal of Accounting Research, 1(4).
Beaver, W.H. McNicjols, M.F. and Rhie, J.W. (2005) “Have Financial Statements Become Less Informative? Evidence from the Ability of Financial Ratios to Predict Bankruptcy”, Review of Accounting Studies, 10,1.
Chava, S, Jarrow, R (2004) “Bankruptcy Prediction with Industry Effects” Review of Finance ,No. 8.
Daubie,M. Meskens, N. (2002) “Business Failure Prediction: A Review and Analysis of the Literature” Working Paper, Department of Productions and Operations Management, Catholic University of Mons, Belgium.
Donoher, W. J (2004) “To File or Not to File? Systemic Incentives, Corporate Control and the Bankruptcy Decision” Journal of Management, 30(2).
Frino, A. Jones, S. Boong Woong, J. (2007) “Market Behaviour Around Bankruptcy Announcements: Evidence from the Australian Stock Exchange” AFAAN Accounting and Finance Vol. 47, No.
Goldberg, D.E (1989) “Genetic Algorithms in Search Optimization and Machine Learning” Publisher Addison- Wesley Longman Publishing Co.
Hossari, G., and Rahman, S, (2005) “A Comprehensive Formal Ranking of the Popularity of Financial Ratios in Multivariate Modelling of Corporate Collapse” Journal of American Academy of Business, No 6.
Henri Hoglond, (2013) “Bankruptcy Prediction Using Support Vector Machine” Journal of Financial Economics, 96, 323–356.
Kennedy, J (1997) “The Particle Swarm: Social Adaptation of Knowledge”, Proceedings of IEEE International Conference on Evolutionary Computation, Indianapolis, IN.
Kennedy, J., Eberhart, R. C (1995) “Particle Swarm Optimization”, in Proc. IEEE International Conference on Neural Networks, IEEE Service Centre, Piscataway, NJ.
Leland , E.H (1994) “Corporate Debt Value, Bond Covenant and Optimal Capital Structure” The Journal of Finance, 49, 4.
Leland, H. E., Toft, K. B. (1996), Optimal Capital Structure Endogenous Bankruptcy and the Term Structure of Credit Spreads, Journal of Finance, 51.
Lensberg, T. Eilifsen, A., McKee, T.E, (2006) “Bankruptcy theory development and classification via genetic programming” European Journal of Operational Research, 169.
Montana, D.J. and Lorence, D. (1989) “Training Feed Forward Neural Networks Using Genetic Algorithms” IJCAI'89 Proceedings of the 11th International Joint Conference on Artificial Intelligence , Volume 1.
Murillo Campello a, John R. Graham B. Campbell R. Harvey (2013) “The Real Effects of Financial Constraints: Evidence from a Financial Crisis” Journal of Financial.
Neilsen, L. Saa_Requejo, J. and Santa Clara, P. (1993) “Default Risk and Interest Rate Risk: The Term Structure of Default Spreads” Working Paper.
Neural Networks: The Case of Bank Failure Predictions”, Management Science, 38(7).

Newton W (1998) “Grant Bankruptcy and Insolvency Taxation” Cumulative Supplement, John Wiley & Sons Inc. New York.

Odom, M., Sharda, R (1990) “A Neural Network Model for Bankruptcy Prediction” Proceedings of the IEEE International Conference on Neural Networks II.

Ooghe, H., De Prijcker, S. (2006), “Failure Processes and Causes of Company Bankruptcy: A Typology”, Department of Accountancy & Corporate Finance, Ghent University, Working Paper, 388.

Ravi Kumar, P., Ravi, V (2007) “Bankruptcy Prediction in Banks and Firms via Statistical and Intelligent Techniques” European Journal of Operational Research, 180, 180.

Rugent, D. Menciuniene, V., Dagilene, L. (2010) “The Importance of Bankruptcy Prediction and Methods” Business: Theory and Practice. Vilnius: Technika, 11.

Shi, Y, Eberhart, R (1998) “A Modified Particle Swarm Optimizer”, in: Proc. 1998 IEEE Intl. Conf. Evolutionary Computation, IEEEWCCI.

Shumway, T (2001) “Forecasting Bankruptcy More Accurately: A Simple Hazard Model”, Journal of Business, 74 (1).

Sprengers M.A (2005) “Bankruptcy Prediction using Classification and Regression Trees” Bachelor Thesis Informatics & Economics Faculty of Economics Erasmus University Rotterdam.

Tam, K.Y., Kiang, M (1992) “Managerial Applications of.

Tamari, M (1966) “Financial Ratios as a Means of Forecasting Bankruptcy” Management International Review 4.

Tan, Y., Xiao, Z.M (2007) “Clonal Particle Swarm Optimization and its Applications”, The 2007 IEEE Congress on Evolutionary Computation.

Turetsky, HF., McEwen, R A (2001) "An Empirical Investigation of Firm Longevity: A Model of the ex-ante predictors of Financial Distress" Review of Quantitative Finance and Accounting, 16.

Wang, J. Meric, G., Lui, Z. (2009) “Stock Market Crashes, Firm Characteristics, and Stock Returns” Journal of Banking and Finance, 33 (9).

Ziegler, A (2004) “A Game Theory Analysis of Options. Berlin” Springer-Verlag, ISBN: 354020668 X.